GUIDELINES FOR OPERATIONAL TEST AND EVALUATION OF SOFTWARE-INTENSIVE SYSTEMS

1. PURPOSE AND APPLICABILITY

DoD increasingly acquires and deploys software-intensive systems[1] as a series of releases within a formal acquisition increment[2], where each successive release builds upon the capability and functionality previously deployed. The capabilities in each release must complete adequate operational test and evaluation before being deployed for use within the Department of Defense.

These guidelines provide a means to tailor and determine "adequate" operational test and evaluation (OT&E) of software-intensive systems using an approach guided by assessment of "risks to mission accomplishment"[3], the ability of the Operational Test Agency (OTA) to assess these risks, the ability to assess improvements provided by each release to the mission capabilities, effectiveness, suitability, and survivability/security of the currently fielded system. These guidelines should be used by the OTA to determine the level of test effort for software-intensive systems and should not be confused with DoDI 8510.XX which must be used to determine cyber security posture risks.

For software-intensive systems on DOT&E oversight, the Operational Test Agencies must obtain DOT&E agreement on the level of risk and the corresponding level of operational testing for all releases that are intended to be deployed.

---

[1] For the purposes of these guidelines, software-intensive systems are computer-based information systems executing one or more resident, separable application software programs. Examples include automated information systems (AIS) and command and control (C2) systems. Software systems embedded in weapon systems are excluded from these procedures. An increment of a software-intensive system is a militarily useful and supportable operational capability that can be effectively defined, developed, tested, deployed, and sustained as an integrated entity or building block of the target system.

[2] For the purposes of these guidelines, an increment is a formal acquisition effort approved by the milestone decision authority. Each increment may have one or more releases constituting a change to the fielded hardware and software baseline.

[3] Risk is a compound function of the likelihood of occurrence, and resulting mission impact, of an increment's failure to be operationally effective, suitable, and survivable. Mission is defined as the objective or task, together with the purpose, which clearly indicates the action to be taken. (DAU Glossary)

These guidelines supersede the Director, Operational Test and Evaluation (DOT&E) memo titled Guidelines for Operational Test and Evaluation of Information and Business Systems (14 September 2010) and are applicable immediately.

2. GENERAL APPROACH

Each formal acquisition increment of a software intensive system will complete at least one full OT&E unless specifically waived by DOT&E. OT&E for each release within a software-intensive systems increment will be guided by an assessment of operational risks to mission success, determination of an adequate level of operational testing, and DOT&E approval.

The lead OTA will complete an initial risk analysis with inputs from the Program Manager, developmental test community, intelligence and counter-intelligence communities, and user representatives to document the probability of occurrence of an adverse event and severity of potential effects on mission accomplishment for the formal acquisition increment. The results of this initial risk analysis are expected to be part of the Test and Evaluation Master Plan (TEMP) for Milestone B and will be used for initial planning of the appropriate level of OT&E, through a tailored approach, to assess operational effectiveness, suitability, and survivability/security[4], e.g. as expressed in the Critical Operational Issues (COIs)**.**

3. LEVELS OF OT&E

Three levels of operational testing are possible for software intensive systems. Programs should plan an integrated test and evaluation strategy to fully assess all capabilities potentially affected by change in a given release. DOT&E and AT&L directives require the seamless integration of developmental and operational testing throughout the life cycle of a system under test. In their joint memo of 25 April 2008 DOT&E and AT&L defined integrated testing as follows:

> Integrated testing is the collaborative planning and collaborative execution of test phases and events to provide shared data in support of independent analysis, evaluation and reporting by all stakeholders particularly the developmental (both contractor and government) and operational test and evaluation communities.

[4] Typically, survivability testing for software-intensive systems will be based on cyber security. In some Services and Agencies, cyber security capability is addressed as security rather than survivability. See also DoDI 8510.XX; CJCSI 6510.01F; the DOT&E policy memo "Test and Evaluation of Information Assurance in Acquisition Programs" dated 01 February 2013; the DOT&E policy memo "Procedures for Operational Test and Evaluation of Information Assurance in Acquisition Programs" dated 21 January 2009; and the DOT&E policy memo "Clarification of Procedures for Operational Test and Evaluation of Information Assurance in Acquisition Programs" dated 04 November 2010.

The design of testing activities at each level of OT&E must be based upon the fundamental objective of evaluating operational effectiveness, suitability, and survivability/security as expressed in the COIs.

Level I T&E

Level I T&E is an evaluation primarily using data from integrated test events other than a dedicated independent operational test event, e.g., developmental tests, certification events, and independent observations of the capability being used in operationally realistic or representative conditions. Level I T&E is appropriate for releases having low risks to mission accomplishment. Typical releases with low risk capabilities where Level I T&E is anticipated are maintenance upgrades, hardware upgrades, and software patches containing only minor capabilities or enhancements.

Features of Level I T&E are:

- The OTA influences and monitors selected test activities including recommending inclusion of test cases for examining specific operational issues, and collecting data for the evaluation.
- Contractor participation is in accordance with the nature of the test events with consideration given to fielding plans of the system and release.
- For acquisition and fielding decisions, the OTA must confirm that the program has plans in place that address recovery from failures and resolution of shortfalls discovered in test events.
- The assessment plan is approved by the lead Service or agency OTA.
- The OTA prepares and provides an appropriate independent evaluation or assessment to support the acquisition and fielding processes and, for programs on DOT&E oversight, provides a copy to DOT&E.

Level II OT&E

Level II OT&E is an evaluation that includes an independent operational event, which is carried out by typical users in an operationally realistic or representative environment to assess specific factors of operational effectiveness, operational suitability, and survivability/security. The evaluation primarily uses data collected during the independent operational event, but also includes data as appropriate from other integrated test program events. Level II OT&E is appropriate for releases having a moderate level of risk with limited potential for mission disruption. The Level II OT&E is typically suitable for modest changes and additions in operational capabilities.

Features of Level II OT&E are:

- Typical users in an operationally realistic or representative environment performing mission tasks. One or more operational sites might participate and the OTA might prescribe scripted events in addition to normal user activity.
- Contractor participation is limited to that prescribed in the program's support plan.

- For system increments intended to be fielded, the OTA must confirm that the program has plans in place that address recovery from failures and resolution of shortfalls discovered in test events.
- Level II OT&E requires completion of the cyber security test process through a cooperative vulnerability evaluation.
- A test concept briefing will be presented to DOT&E 180 days prior to start of Level II OT&E.
- DOT&E will approve the operational test plan or equivalent document, which should be submitted to DOT&E at least 60 days prior to the start of Level II OT&E, with approval required before start of Level II OT&E.
- All test data will be provided to DOT&E for independent analysis and reporting.
- The OTA prepares an independent evaluation of operational effectiveness, operational suitability, and survivability/security to support the acquisition and fielding processes and provides a copy to DOT&E.

Level III OT&E

Level III OT&E is an end-to-end evaluation of the operational effectiveness, operational suitability, and survivability/security of the operational capability using the COIs and an independent dedicated operational test. Level III OT&E is the highest level and most comprehensive of OT&E, and is appropriate for significant or new operational capabilities with high risk of mission disruption.

Features of Level III OT&E are:

- Level III OT&E must comply with statutes and all provisions of the DoD 5000 series regulations.
- The OTA carries out test events in an operational environment.
- Level III OT&E requires completion of the cyber security test process through independent threat representative penetration testing.
- A test concept briefing will be presented to DOT&E 180 days prior to start of dedicated OT&E.
- DOT&E will approve the operational test plan, which should be submitted to DOT&E at least 60 days prior to the start of dedicated OT&E, with approval required before start of dedicated OT&E.
- All test data will be provided to DOT&E for independent analysis and reporting.
- The OTA independently evaluates and reports on the operational effectiveness, operational suitability, and survivability/security using all available data, especially independently collected operational test data, to support the acquisition and fielding processes with a copy provided to DOT&E.

4. IMPLEMENTATION

Tailoring the OT&E for each release within a software-intensive system formal acquisition increment will follow a three-step process of assessing risks to mission accomplishment, determining an adequate level of operational testing, and obtaining DOT&E approval. The basic method is to assess the level of risk (likelihood and mission impact of an adverse event) of the operational capabilities comprising the release. This assessment determines an appropriate level of OT&E for the operational capabilities and thus the release. Some capabilities may be assessed using data only from integrated testing, while within the same release an independent operational test may be needed to assess other more critical capabilities, capabilities most affected by change, or where risk is dependent on the operator actions or operational environment.

The entire risk assessment and design/conduct of testing process should be a significant focus area for continuous improvement. Encountering significant risks after completion of testing indicates that the risk assessment process, operational test adequacy, and the test/fix/test process require improvement. A simple software reliability metric of the previous software release should be shown as part of the risk assessment and level of test package when submitted to DOT&E for approval. Figure 1 is a notional chart showing the cumulative number of priority 1 and 2 problems[5] encountered, and cumulative priority 1 and 2 problems fixed or downgraded due to discovery of operationally acceptable workarounds, starting from government developmental testing and continuing through operational testing and extending to include all usage of the software after completion of operational testing. Test periods are also shown on the chart.

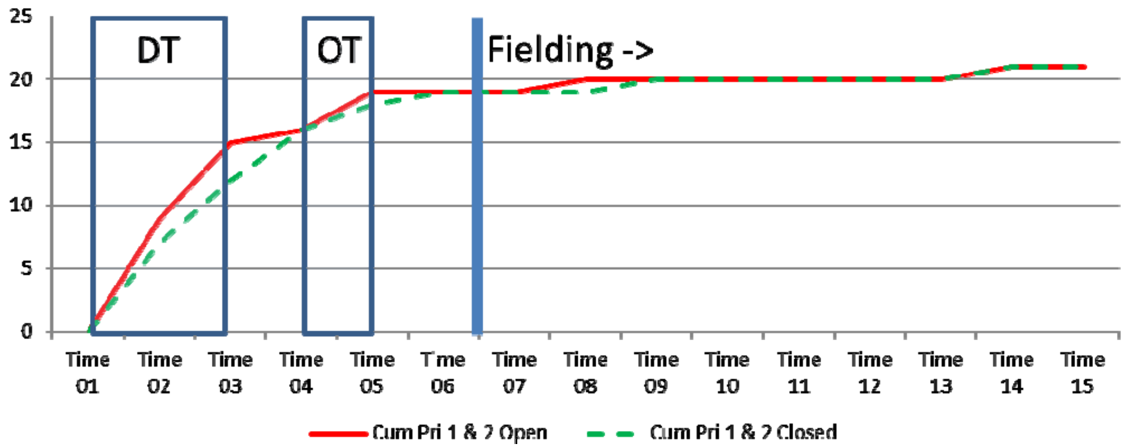

Figure 1. Notional Metric Chart

Amelioration of risk will be addressed by planning for sufficient time for fix and regression testing in schedules, and by planning to test operational workarounds early and with user feedback to ensure these issues are correctly rated as to mission impact and that the workarounds are operationally acceptable, both individually as well as in the context of all operational workarounds affecting system usage.

[5] As defined in IEEE/EIA Standard 12207.2-1997, Annex J.

Risk Assessment

The risk assessment combines the distinct concepts of identifying events that are considered as adverse within the system under test, assessing likelihood of occurrence of these events, and the likely operational effects of a capability failing to be operationally effective, suitable, and survivable/secure. Several methods and approaches are possible for completing the risk assessment. This document provides guidance on methods along with questions and considerations for performing the risk assessment.

The OTA, with support from the program management office, user representative, developmental test organization, and intelligence and counter-intelligence communities, assesses and documents the risks. Risk assessments are developed using the OTA's preferred procedures. Assessments must distinguish between the likelihood of occurrence and the severity of the mission impact if the risk is realized (no matter how unlikely). The OTAs may have or develop their own risk rating scales. A three-level scale is used in this discussion for illustrative purposes.

Identify Risks

Risk assessment begins with Risk Identification[6]. DOT&E expects the OTAs to evaluate risk categories, questions, and considerations that best reflect the release and operational capabilities being assessed. The focus is to identify and assess significant risks to operational mission success that might occur once the system is deployed, not the technology maturation and manufacturing risks that might prevent system acquisition. Risk to operational mission success may be divided into four primary categories (as shown below in solid bullets), and within these primary categories, there are several important sub-categories of events that could impact system operations and mission accomplishment. Four risk categories are:

- Technology and Software Development (including software reliability). This risk category represents the well-known concern that software can have “bugs” and/or be developed with incorrect understanding of user needs or the operational environment.
- Integration and Deployment. This risk category relates to signal and data environment; program interfaces to the operating system and user input, interfaces to legacy databases, messaging, communications protocols, and local configuration files; published and actual software service specifications; interoperability; real time processing issues; competency and accurate record-keeping of system administrators tasked with software installation, and other aspects of distributed computing.
- Training, Utilization, and Management. This risk category relates to user training and organizational buy-in; tactics and procedures for sustainment; and usability issues.
- Cyber Security/Information Assurance. This risk category relates specifically to the survivability/security assessment. See footnote 4.

[6] See also the information assurance risk assessment methodology in NIST SP 800-30 or the Software Engineering Institute's "Taxonomy-Based Risk Identification," CMU/SEI-93-TR-6 ESC-TR-93-l83.

This list is not all-inclusive, but rather intends to convey some of the most common risks encountered over years of testing such systems. Various types of testing may be best suited for addressing each specific risk area. An iterative development/test/fix cycle used throughout developmental testing cycle will help provide additional insight into the driving factors and how to best test them. Tracking of test completion of use cases during developmental and operational testing will contribute to the reliability program.

The risk categories include human and organizational factors that are involved in deployment, training, and business process change. Operational deployment, training, and process issues may be more complex than technical issues in software-intensive projects and may lead to significant mission shortfalls. Human and organizational factors must also be considered in risk assessments and OT&E. Appendix A provides additional guidance for assessing risks in the four general risk categories mentioned above.

Likelihood of Risk Occurrence

Once events that could adversely impact operations have been identified, the risk assessment process will estimate the likelihood that the events will occur. The OTAs may use any scale that they commonly use for determining likelihood of event occurrence, while Table 1 gives an example three-level scale. When in doubt, the likelihood should be rated high. Adjustments to likelihood of occurrence estimates may be warranted as testing progresses through the developmental test/fix/test process.

Table 1. Likelihood Estimate Levels

Estimate of Likelihood of Event Occurrence during operations, given the program's demonstrated maturity to date		
Level	Descriptor	
1	Negligible	One can reasonably assume no occurrence
2	Possible	Issue is possible, but unlikely. Issue cannot be ruled out.
3	Likely	Issue has a significant chance of occurrence. Occurrence would not be surprising.

Operational Impact from Event Occurrence

The assessment of *operational* impact, which is the operational consequence from the event occurring, is somewhat different from the assessment of impact in a standard risk assessment[7]. First, operational impacts relate only to performance effects, not effects on cost and schedule. Second, realization of some events can cause performance effects that do not greatly affect the operational mission, and therefore have low operational impact. For example, redundant capabilities or systems could be used to accomplish that portion of the mission. Thus, operational impact involves an understanding of performance effects from

[7] Such as defined in the "Risk Management Guide for DoD Acquisition" (http://www.dau.mil/pubs/gdbks/docs/RMG%206Ed%20Aug06.pdf).

events happening plus an assessment of the operational mission relevance of those performance effects.

The operational impact question is: If this event occurs and affects the performance of the capability, will that performance effect undermine mission goals?

In order to determine operational impact, the risk assessment must first identify associated performance effects. Software capabilities can fail (hang or crash a system), store or transmit incorrect data, emit confusing messages and graphics, add new avenues of unauthorized access, slow system response, hinder training, and so on. The risk assessment must provide the specifics of how a realized risk would unfold.

After assessing how the anticipated risks to a capability might unfold as performance effects, the OTA must determine how the performance effects could translate into operational impacts on the mission goals. Mission goals are expressed in the measures of operational effectiveness, operational suitability, and survivability/security. The OTAs should use their preferred risk assessment approaches to assess operational impact for each risk/performance effect/mission goal combination. Table 2 is an example of a three-level scale for operational impacts.

Table 2. Operational Impact Levels

Operational Impact Level	Descriptor	Definition
1	Minimal	Annoying system characteristic or nuisance that does not degrade operational/mission effectiveness, suitability, or survivability/security. Little to no impact on mission critical capability.
2	Moderate	Performance effect degrades operational mission effectiveness, suitability, or survivability/security, and no acceptable operator compensation or workaround exists. Performance effect prevents operational mission performance, but can be overcome with operator compensation/workaround. Mission critical capabilities are moderately dependent upon increment performance.
3	Severe or Catastrophic	Performance effect prevents achieving operational mission effectiveness, suitability, or survivability/security threshold, and no workarounds are available. The capability is required for mission success, and its malfunction could cause significant damage to the installed system, to other interconnected systems, or to personnel. Mission critical capabilities are critically dependent upon the increment performance.

Determine Required Level of Operational Test

On completion of the risk analysis effort, the level of OT&E for the operational capabilities and the release can be determined using the likelihood of event occurrence and operational/mission impact for the risks. Table 4 is an example for determining level of OT&E for the example three-level scales of Tables 2 and 3. The required level of OT&E for each capability is the maximum of the OT&E levels determined for each of the risks. The required level of OT&E for the release is likewise the maximum of the OT&E levels determined for each of the capabilities.

Table 2. Level of OT&E Required

Likelihood of Event Occurrence (before test)	3	II	III	III
	2	I	II	III
	1	I	I	II
		1	2	3
	Operational/Mission Impact of Event			
I = Level I OT&E, II = Level II OT&E, III = Level III OT&E				

The determination of required level of OT&E is just the start of the process of actually designing an adequate operational test. Design of an adequate operational test must consider the risks inherent with the system prior to operational testing, the ability of the testing process to manifest the adverse events identified, the ability of the testers to recognize the adverse events once manifested, and the ability of the test/fix/test process to identify, fix, and verify corrections prior to fielding. Quality development processes, developmental test adequacy, and operational test adequacy are all critical to ensuring software intensive systems support mission accomplishment. As stated earlier, software maturity metrics will be used to indicate the need for significant improvement in activities relating to risk assessments and operational test adequacy.

Obtain DOT&E Approval

Once the risk assessment is complete, for software-intensive systems on DOT&E oversight, the OTA will provide DOT&E with the risk assessment (likelihoods of occurrence and mission impacts) and the corresponding proposed level of OT&E for approval. A test concept briefing is used to obtain DOT&E approval for proposed Level II or Level III operational testing.

5. POINT OF CONTACT.

The DOT&E Deputy Director for Net-Centric and Space Systems is the point of contact for these procedures. Issues pertaining to implementation of these procedures for a specific software-intensive system on DOT&E oversight may be addressed to the Deputy Director responsible for the applicable system.

APPENDIX A

RISK ASSESSMENT CATEGORIES AND GUIDANCE

The following outline, with some example questions, represents topics that should be addressed when assessing the four risk categories. Questions in the risk identification phase should be tailored as appropriate for the particular software intensive system that is being assessed. The OTAs are encouraged to add their own questions and any additional risk categories they think appropriate.

- Technology and Software Development
    - Requirements
        - DT testing should ensure that the product meets the stated requirements from a functional perspective. Early DT-level test cases should focus on both breadth and depth of testing in those areas where requirements mismatches are suspected.
        - Third-party systems, when incorporated as part of a larger system of systems, are generally developed with their own requirements contained in their program's requirements documents or other requirements management system. These requirements may not totally align with the requirements of the system under test, either from a functional point of view, or from a system performance and stress point of view.
        - Do the requirements documents clearly and unambiguously state the performance requirements, testability metrics, use cases, and operational constraints for the software? Have mission needs been adequately described and user requirements clearly identified? Is the capability traceable to requirements? Do the requirements address operational needs rather than specifying a technical solution?
        - How stable were/are the system requirements?
    - Test pedigree

- Third-party systems are generally tested against their own requirements, with their primary user base, data sets, and concepts of operation that may not accurately reflect how these third party systems would be used in the operational environment of the system under test. An adequate test for one operational environment and usage may not be adequate for that system being employed in another environment, under a different set of stressors.
- If the capability is primarily commercial off-the-shelf (COTS), non-developmental item (NDI), or government off-the-shelf (GOTS), what is the past performance and reliability? For new technologies, what is the performance record in other applications? Are custom modifications of base software planned?

- Functional software failure
  - Risk of functional software failure should be carefully tracked through early testing by metrics programs showing both test thoroughness and cumulative finding and fixing of problems. These metrics should be a key part of the reliability growth strategy for software intensive systems.
  - Is there a plan for collecting and reporting software reliability metrics, including discovery of new failure modes and closure rate? Does the metrics plan include measures addressing test case completion and thoroughness? What do the metrics actually show?
  - Does the capability or system present any safety hazards to the operators or operational environment? What data and physical effectors does it control and what damage can they do? Could users of the system unknowingly use bad information from the system to make bad judgments, leading to serious mission or safety consequences?
- Software complexity
  - How complex is the capability (lines of code or other industry standard complexity metrics)?
  - How dependent is the capability upon new technologies (hardware and software) such as virtual servers, information services, or agile development methodologies?

    - What is the commercial tempo of change in the technology areas represented in the capability and how mature are the technologies that are used?
    - How many agents (government, contractors, sub-contractors) participated in the development of this capability?
    - What is the proportional change to system hardware and software introduced by the new capability (100 percent new; small patch to large deployed system; etc.)?
    - What is the cumulative change to system hardware and software since the last full operational test?
    - Does the capability implement a change in executive software (operating system or database management system)?
    - Does the new capability introduce changes that place in jeopardy or modify the system data structures?
    - Does the capability introduce any new standards or protocols?
    - When errors are manifested in the software, are they easily recognized by the test community?
- Server performance
    - Major risks to server performance may involve the time required to process large amounts of data or operating system ability to manage application tasks. Performance monitoring should be used to help characterize and reduce risk related to server performance. As servers are more heavily used, performance should degrade gracefully.
- Client performance
    - Risks with clients may include conflicts in integration or memory leaks where the client performance may slowly degrade over time, eventually causing the client to be rebooted. Web browser incompatibilities should also be considered.
- Network performance
    - Risks to network performance should consider available bandwidth, network delays, amounts of data required to be passed over the network, network overhead associated with transferring data, and circuit path reliability. Aspects of network performance should include not only the local area networks at the operational

locations, but also to all critical interfacing nodes and data sources. Performance monitoring systems can be used to monitor network traffic and to identify software problems manifested through poor network performance.

- Developer environment and track record
    - This issue concerns how well the developer's test environment can match operational realism, to include hardware, networks (including network delays, loading, and bandwidth), and whether the developer has realistic operational data with which to test. The ability of the developer's test team to thoroughly understand the user data, user processes, and user mission are critical. Early user involvement can be used to help mitigate risk in this area. Developer environment is a particular area in which past track record should be considered. Early developmental testing at a government DT facility can also help reduce risk.
    - For newly developed software, what is the developer's Capability Maturity Model (CMM) rating as defined by the Software Engineering Institute? Do cost or schedule pressures adversely affect the ability of the developer to perform to their CMM rating?
    - Were any Priority l or Priority 2 problems experienced with previous increments from this development team?
    - Does the developer employ a robust set of software management indicators?
    - Does the developer's environment allow the developer to replicate problems found in the field or in government test facilities?
    - Does the developing contractor's test agent have sufficient experience and technical expertise to conduct a proper technical evaluation? Was thorough integration and regression testing conducted? Have clear exit criteria been identified for developmental testing of this capability? If a problem is manifested in the development environment, can the test agent recognize it as a problem?

- Integration and Deployment
    - Interoperability problems
        - Interoperability issues can arise when third party systems may not develop their software to exactly meet information exchange

requirements (IERs), or may not adequately test their systems for compliance to the IER. Live test limitations can lead to the use of test data feeds and test databases that may not reflect operational reality with respect to data errors. System configuration issues and software version control issues can also be high risk aspects of interoperability.

    - Does the capability require integration with another new or immature system? Are interfaces with existing systems fully documented and under configuration control?
    - How complex are the external system interface changes (hardware, software, data, signals, and messaging) required for the capability?
    - Are mature Interface Control Documents available and approved?
    - Must the capability interoperate with other systems? Are any of those systems also in development?

- Data flow issues (i.e. through mission threads)
    - Data flow issues through mission threads tend to be due to integration, interoperability, or system configuration issues. The OTA should consider the thoroughness and track record of the DT test program to satisfactorily address these issues.
- System administration challenges
    - Specific challenges include the time and complexity of building and configuring the system. Have they been afforded sufficient time to do this?
    - How complex is the build/configuration process, and how accurate is the documentation they work from?
    - How difficult is the system password change process?
    - Does the integration of the entire system (e.g., hardware, software, communications, facilities, management, operations, sustainment, personnel) present unusual challenges?
- Test location build, integration, and configuration
    - If testing in other than an operational location, how similar are the environments, the build/configuration processes, and what is the extent of any site-unique differences?
    - Does the test location accurately reflect the starting position that will be encountered during operational fielding?

  - Operational site build, configuration, and operational cutover
    - What is the extent of any site-unique differences?
    - Are there any coalition issues that may not have been tested?
    - Does the site have the ability to continue operations while the build/configuration process is on-going? Has the site developed plans for continuity of operations during the installation of the new increment?
    - Does the site have the means, and a plan, to thoroughly check out mission thread completion before operational cutover?
  - Data errors in databases or streaming data
    - Data errors can occur when software used to input values is not sufficiently robust. Do operational databases contain known data errors or data not in strict compliance with specifications (to include blank fields)?
    - Must the capability interact with fielded, legacy databases?
    - Must the capability interact with systems that produce streaming data? Do input data streams or message traffic sometimes not follow strict standards?
    - How complex are the user interactions? How mature is the interface that captures and conditions the user input?
- Training, Utilization, and Management
  - Ops Tempo and system stress
    - Does real world ops tempo prevent enough users from participating in testing activities?
    - Does the program and test community have an automated means of inducing system stress that is operationally realistic? Has it been verified and validated?
    - Have aspects of ops tempo such as shift change-over and collaboration been considered?
    - Has the program been stressed to, or beyond, threshold concerning number of users, operational tempo, data flow rates, and contention/throughput on networks? Are system logs monitored to ensure sufficient system capacity is maintained?
  - Training and skills retention

- Have the users been trained on the system? Have the users had sufficient practice using the system to accomplish their mission?
- When was the training conducted, and has it grown stale?
- Is there a sufficient cadre of highly experienced users to mentor lesser-experienced users?
- Do the operators possess the skill levels required to use the increment's capabilities effectively? Has an adequate training plan been developed or implemented to include reorientation and sustainment training, as well as incorporating changes to new user training?

o Lack of robustness

- Have the human factors and man-machine interface impact on the system performance been adequately considered?
- Is it too easy for novice users to make mistakes when entering data? Does the system effectively trap user input errors and notify or help the user to fix the input?

o Cumulative effects of operational workarounds and high priority problems

- High priority problems may have workarounds that are either difficult for users or system administrators to remember to execute, or else there may be so many workarounds that the overall burden to the user is simply overwhelming. What is the expected effect of all known mission-impacting workarounds?

o Documentation

- A significant risk in the area of documentation lies with manuals used by system administrators to build, configure, and maintain the system. What is the state of system build/configuration documentation regarding document red-lines?
- Does system documentation assist with addressing site-unique differences from the build/configuration of the test article?

o Help Desk

- Can the help desk process maintain support to daily operations while assisting with the site build/configuration/checkout process?
- Are all tiers of help desk properly trained and equipped to address problems during fielding and early system use?

- Ability of program office to respond to system problems
  - Does the software reliability metrics program track time to fix high priority problems? What does this reveal?
  - What is the track record of the program office effectively addressing high priority problems found in the field? Note that being able to do this well is a good thing, while having to do it frequently may not be.
- Change Management, Commitment, CONOPs, TTPs
  - Is the user committed to the successful implementation of the new capability? Is the receiving organization committed to the successful implementation of the new capability?
  - How extensively have prototypes been used to evaluate acceptance by typical users? Is the user interface intuitive or does it require extensive explanation?
  - Have operational and user support procedures been developed and readied for implementation? Have user representatives developed appropriate concepts of operations, policies, procedures, training, support, and contingency plans for a full operational deployment?
  - Is the receiving organization prepared for the changes in business processes or TTPs associated with the new capability?

- Information Assurance
  - The IA portion of the risk assessment will be conducted according to DOT&E Memo, Policy for Operational Testing of Information Assurance in Acquisition Programs, 21 Jan 2009 and the DOT&E Memo, Clarification of Procedures for Operational Test and Evaluation of Information Assurance in Acquisition Programs, 4 Nov 2010.
  - The OTA should use all available IA-related data from certification, accreditation, and developmental testing to assess the risk of the system before entering operational testing. IA-related data pertaining to operational use of the legacy system may also be relevant.
  - The IA risk assessment should include aspects of where the system sits in a network, relative to security protection and detection mechanisms.
  - The IA risk assessment may also consider whether the aspect of “React” includes the ability to conduct sufficient forensics to determine scope of

damage or information loss to an adversary, and if the element of surprise has been lost.

- The aspect of "Restore/COOP" will also consider the ability to sustain mission operations under conditions where the system may be partially or totally inoperable for an undetermined period of time. The IA assessment for "Restore/COOP" should also consider overall user confidence and trust in the system and data.
- Additional IA questions could include:
  - What is the status of the Certification and Accreditation package?
  - For third party systems, do they require separate Authority to Operate (ATO), and if so, status of these?
  - Are there foreign sources of component chips? Is there a risk of counterfeit parts?
  - Are the network interfaces and information exchanges defined, including all relevant data sources?
  - Does the new capability affect system security via new or altered external interfaces?
  - Is administrative access granted to software in order to enable installation? Are new high-access accounts required for the newly installed capabilities? Are all system passwords changed after installation, and is this process relatively easy and error-free?
  - Do security protocols of new deliverable map cleanly to existing protocols?
  - What is mission assurance capability (MAC) and confidentiality level (CL) for the deliverable?
  - Have the DoD 8500 information assurance controls been assessed?
  - Who will do the scanning for the vulnerability assessment?
  - Who will do the vulnerability assessment and penetration testing if necessary?
  - Who will identify mitigation techniques for vulnerabilities found?
  - What is the current state of known IA vulnerabilities in the fielded system?
  - Are there tools available to detect penetrations?

- If vulnerabilities are detected, are the react and respond procedures identified?
- Supply chain threat and lifecycle threat?
- When is the last time the system was actively aggressed by a Red Team, at what threat level, and what was the outcome?